JOURNAL OF VIROLOGY, May 2005, p. 5249–5258 Vol. 79, No. 9
0022-538X/05/$08.00+0 doi:10.1128/JVI.79.9.5249–5258.2005

The Common Viral Insertion Site *Evi12* Is Located in the 5′-Noncoding Region of *Gnn*, a Novel Gene with Enhanced Expression in Two Subclasses of Human Acute Myeloid Leukemia

Eric van den Akker,* Yolanda Vankan-Berkhoudt, Peter J. M. Valk, Bob Löwenberg, and Ruud Delwel

Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

Received 19 July 2004/Accepted 22 December 2004

The leukemia and lymphoma disease locus *Evi12* was mapped to the noncoding region of a novel gene, *Gnn* (named for *Grp94* neighboring nucleotidase), that is located immediately upstream of the *Grp94/Tra1* gene on mouse chromosome 10. The *Gnn* gene is conserved in mice and humans. Expression of fusion constructs between *GFP* and *Gnn* cDNA isoforms in HEK-293 cells showed that Gnn proteins are located mainly in the cytoplasm. Immunoblotting experiments demonstrated the presence of multiple Gnn protein isoforms in most organs, with the lowest levels of expression of the protein detected in bone marrow and spleen. The *Evi12*-containing leukemia cell line NFS107 showed high levels of expression of a ~150-kDa Gnn isoform (Gnn^{107}) that was not observed in control cell lines. Overexpression may be due to the viral insertion in *Evi12*. The Gnn^{107} protein is probably encoded by a *Gnn* cDNA isoform that is expressed exclusively in NFS107 cells and that includes sequences of TU12B1-TY, a putative protein with homology to 5′-nucleotidase enzymes. Interestingly, using Affymetrix gene expression data of a cohort of 285 patients with acute myeloid leukemia (AML), we found that *GNN/TU12B1-TY* expression was specifically increased in two AML clusters. One cluster consisted of all AML patients with a t(8;21) translocation, and the second cluster consisted of AML patients with a normal karyotype carrying a *FLT3* internal tandem duplication. These findings suggest that we identified a novel proto-oncogene that may be causally linked to certain types of human leukemia.

Cloning of viral integration sites from retrovirally induced mouse hematopoietic malignancies has resulted in the identification of many leukemia disease loci. Proviral insertions activate proto-oncogenes or inactivate tumor suppressor genes, thereby interfering with normal hematopoiesis, which ultimately leads to leukemia (for a review, see reference 7). The application of improved reverse transcription-PCR (RT-PCR) and inverse PCR strategies together with the availability of mouse genome databases has accelerated the identification of common virus integration sites (cVISs) (9, 11, 13, 17).

We recently identified a novel cVIS, *Evi12*, in leukemias and lymphomas induced by Cas-Br-M murine leukemia virus (8, 20). Southern blot analysis showed that rearrangements were present in 14% of tumors and that in each case only one allele was affected, strongly suggesting that the viral target gene is a proto-oncogene. All proviral insertions occurred in a 1.7-kb region upstream of the molecular chaperone gene *Grp94/Tra1* on mouse chromosome 10 (20). Interestingly, others have described the same locus as a common viral target locus in retrovirally induced lymphomas in AKXD mice (17), providing additional evidence that *Evi12* plays an important role in the development of leukemia and lymphoma. From our previous experiments, we concluded that overexpression of *Grp94/Tra1* was unlikely to be the cause of transformation: Grp94/Tra1 is a chaperone protein ubiquitously expressed in the endoplasmic reticulum, and no differences in *Grp94/Tra1* mRNA or protein expression were observed between leukemias with or without integration in *Evi12* (20).

The aim of this study was to further characterize the genomic area encompassing the *Evi12* locus in order to find the gene affected by retroviral insertion in *Evi12*. Here we report on the identification and characterization of a novel gene, *Gnn* (named for *Grp94* neighboring nucleotidase), that is located immediately upstream of *Grp94/Tra1*. *Gnn* is abnormally expressed in a murine leukemia cell line harboring a viral insertion in *Evi12*, and strikingly, two subtypes of human primary acute myeloid leukemia (AML) show enhanced *GNN* expression.

MATERIALS AND METHODS

Exon trapping. Exon trapping was performed as described earlier (19). Briefly, a bacterial artificial chromosome clone encompassing 150 kb of the *Evi12* locus was partially digested with HpaII. Fragments were cloned into the exon trap vector pERVF0, pooled, and transfected into COS cells. RNA was isolated after 2 or 3 days and used to amplify potential exons by RT-PCR. Southern blot analysis confirmed the presence of one of the isolated potential exons (*Gnn* exon 3) on a 2.8-kb EcoRI/BamHI genomic fragment near *Evi12*.

Synthesis of cDNA. Total RNA was extracted from cell lines and adult organs using guanidinium isothiocyanate or Trizol (Gibco BRL, Breda, The Netherlands). Reverse transcriptase reactions were performed with 5 μg of RNA using random hexamers, oligo(dT), or oligo(dT) adapter primer [5′-GTCGCGAATT CGTCGACGCG(dT)$_{15}$-3′] for 3′ rapid amplification of cDNA ends (3′ RACE) using the superscript cDNA amplification kit (Gibco BRL).

Identification of *Gnn* sequences. A mouse 17-day Embryo MATCHMAKER (MM) cDNA library (BD Biosciences, Palo Alto, Calif.) was initially used to amplify additional *Gnn* sequences. The locations of amplified *Gnn* cDNA frag-

* Corresponding author. Mailing address: Department of Hematology, Erasmus University Medical Center, Dr. Molewaterplein 50, 3015 GE Rotterdam, The Netherlands. Phone: 31 10 4087034. Fax: 31 10 4089470. E-mail: h.vandenakker@erasmusmc.nl.

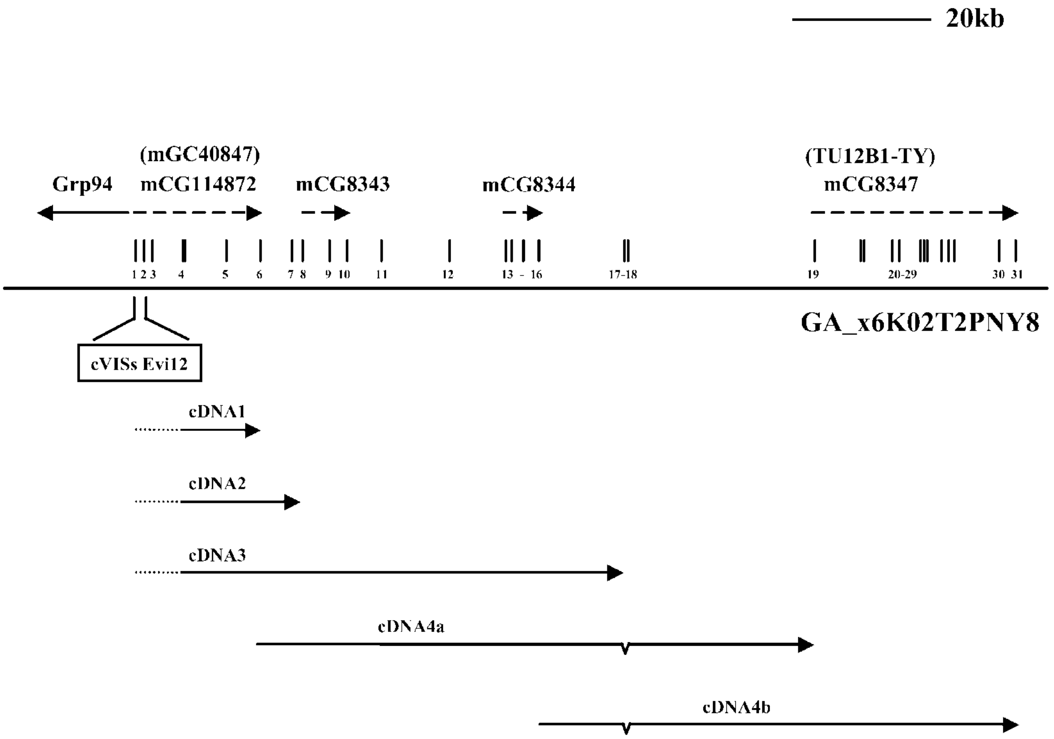


FIG. 1. Overview of the *Evi12*/*Gnn* locus on mouse chromosome 10. *Evi12* is located upstream of *Grp94* between the first exon and second intron of the novel *Gnn* gene. The locations of the 31 identified *Gnn* exons (short vertical lines above the map) are indicated, as well as the exons included in *Gnn* cDNA1 to cDNA4 (arrows). Putative Celera genes overlapping with the *Gnn* sequence are indicated (dashed arrows above the map). *Gnn* exon 3 was identified using the exon trapping strategy. The translation initiation site of *Gnn* is located in exon 4. The complete coding regions of *Gnn* cDNA1 to cDNA3 were amplified using primers located before the start in exon 4 and in their respective 3′ UTRs. We found that the 5′ UTRs of these cDNAs (dotted parts of cDNA arrows) may vary in the inclusion of exons 2 and 3. *Gnn* cDNA4a and cDNA4b are partial sequences that overlap and that were amplified using gene-specific primers for exons 6 and 19 and exons 16 and 31, respectively. The exclusion of *Gnn* exons 17 and 18 in cDNA4a or cDNA4b is indicated. The complete sequence of cDNA4 is likely to be a composite of the sequences overlapping in cDNA3, cDNA4a, and cDNA4b.

ments cDNA1 to cDNA4 are indicated in Fig. 1. The sequences of cDNA1 were amplified by PCR using primers for exon 3: 5′ sequences were amplified using pACT2MM-5′ (BD Biosciences) and 5′-ACCTCCTCATGGTCTGTGGG-3′ and then pACT2MM-5′ and 5′-GGTAAAAGGGCCATATCTTC-3′, and 3′ sequences were amplified using pACT2MM-3′ (BD Biosciences) and 5′-GAAGATATGGCCCTTTTACCC-3′ and then pACT2MM-3′ and 5′-CACAGACCATGAGGAGGT-3′. The full-length coding region of cDNA1 was amplified from the embryonic library using EcoRI-tagged primers for exon 4 (5′-GCCGAATTCCATCCTGGAGCCGAGTGAA-3′) and exon 6 (5′-GACGAATTCTTCAGAGAGTCTAGCAGGGG-3′).

3′ RACE on cDNA from NFS107 cells using a primer set for exon 6 resulted in the identification of exons 7 and 8 as they are found in cDNA2. Primers for the first reaction were 5′-GCTCTTGTCTGGGGAACG-3′ and adapter primer, followed by 5′-GCTGTGGGAAGTCCACAC-3′ and adapter primer. The coding region of cDNA2 was amplified with primers for exon 4 and exon 8 using cDNA from NFS107 cells. For the primary PCR, the primers were 5′-CCTTGGCGACTGGGCCAGG-3′ and 5′-GCACCAGCAGGGGGCAGC-3′. For the nested PCR, the following primers were used: 5′-GCGACATCATCCTGGAGCC-3′ and 5′-ATCACACACCTGGAATCACGG-3′.

RT-PCRs on cDNA from NFS107 cells using primers for the coding region of Celera gene mCG8344, which is located downstream of *Evi12* and cDNA1/2, resulted in identification of the 3′ end of cDNA3 (part of exon 16 to the end of exon 18). The primers were 5′-CAAGGAGAAGGCAGATACG-3′ and the adapter primer for the primary PCR and 5′-CCTCAGCGACACCTTGTGT-3′ and the adapter for the nested PCR. Subsequently, the complete cDNA3 coding sequence was amplified with primers for exon 4 and exon 18. The primary PCR was performed with 5′-CCTTGGCGACTGGGCCAGG-3′ and 5′-TTTATTCCTTATTGGGGTCATC-3′, and the nested PCR was performed with 5′-GCGACATCATCCTGGAGCC-3′ and 5′-TTTATTCCTTATTGGGGTCATC-3′. The *Gnn*/*TU12B1-TY* fusion cDNA4a was amplified from cDNA of NFS107 cells using primers for exon 6 and exon 19. The primary PCR was performed with 5′-GCTCTTGTCTGGGGAACG-3′ and 5′-GATGTCTGACAGGCTCATC-3′, and the nested PCR was performed with 5′-GCTGTGGGAAGTCCACAC-3′ and 5′-GTTCATGATGGATGGAACC-3′. *Gnn*/*TU12B1-TY* exons 19 to 31 (cDNA4b) were identified by RT-PCR using primers for exon 16 and in the last exon of the Celera mCG8347 gene. The primary PCR was performed with 5′-CAAGGAGAAGGCAGATACG-3′ and 5′-GGATTCTGGTCTGTTCGG-3′, and the nested PCR was performed with 5′-CCTCAGCGACACCTTGTG-3′ and 5′-CAAGCTCCCAAACTGGGC-3′.

To compare the expression of the *Gnn*/*TU12B1-TY* fusion and other *Gnn* products between NFS107 and control cell lines, we performed the RT-PCR that amplifies *Gnn* cDNA4a, and control PCRs that amplify the 3′ end of *Gnn* cDNA1 (exon 6) and the 3′ end (exons 16 to 18) of *Gnn* cDNA3 (see above).

All reactions with one exception were performed with Expand polymerase (Roche, Mannheim, Germany) under the following conditions: 30 cycles, with 1 cycle consisting of 1 min at 94°C, 1 min at 58°C, and 3 min at 72°C. The exception was amplification of the cDNA1 coding region, which was performed with *Pfu* polymerase (Promega, Madison, Wis.). The conditions for this reaction were as follows: (i) 3 cycles, with 1 cycle consisting of 1 min at 96°C, 1 min at 55°C, and 3 min at 72°C; (ii) 27 cycles, with 1 cycle consisting of 1 min at 96°C, 1 min at 60°C, and 3 min at 72°C.

Cloning of expression constructs. The PCR product of cDNA1 was digested with EcoRI and either directly inserted into the EcoRI site of enhanced green fluorescent protein (EGFP)-C1 (BD Biosciences) to generate a construct encoding EGFP-Gnn1 or blunted and inserted in HpaI-digested PLNCX (BD Biosciences) to generate a nontagged construct encoding Gnn1. PCR products of cDNA2 and cDNA3 were first ligated into the TA cloning vector (Invitrogen, Breda, The Netherlands). The integrity of all three cloned PCR products was checked by sequencing. The correct clone of *Gnn* cDNA2 was excised with EcoRI and inserted into the EcoRI site of EGFP-C1 to generate a construct encoding EGFP-Gnn2. PLNCX-Gnn2 was generated by first blunting the EcoRI-excised *Gnn* cDNA2 and then inserting it in HpaI-digested PLNCX. All analyzed clones of *Gnn* cDNA3 in the TA cloning vector contained one or more single-nucleotide PCR errors, but each clone had a different error. One clone had an error in a region overlapping cDNA2. To generate the correct EGFP-cDNA3 fusion, TA-cDNA2 was digested with EcoRI and SstI, and TA-cDNA3 was partially digested with SstI and EcoRV. The resulting fragments were ligated simultaneously into EcoRI/SmaI-digested EGFP-C1 to generate EGFP-Gnn3. To generate PLNCX-Gnn3, the two fragments were simultaneously ligated into HpaI-digested PLNCX.

Sequence analysis. PCR products cloned in the TA vector were sequenced using an ABI 3100 sequencer (Perkin Elmer, Nieuwerkerk a/d IJssel, The Netherlands) with forward and reverse primers (from the TA cloning kit) and *Gnn*-specific primers.

Transfection and immunofluorescence. HEK-293 cells were grown on glass coverslips and transfected by the calcium phosphate transfection method. Two days after transfection, cells were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) at 4°C for 20 min. Fixed cells were washed three times with PBS, embedded in Vectashield (Vector Laboratories, Burlingame, Calif.), and observed by confocal laser scanning microscopy.

Generation of Gnn antibody. An affinity-purified rabbit antiserum against Gnn was prepared by Research Genetics (Huntsville, Ala.). The antiserum was raised against the N-terminal DFQEERDFLAKSIFPNLD sequence of Gnn, which is 100% conserved in mice and humans.

Immunoblotting. Cell pellets or homogenized tissues were lysed in ice-cold lysis buffer (50 mM Tris-HCl [pH 8], 100 mM NaCl, 1% Triton X-100, 1% Pefabloc SC, 50 μg of aprotinin/ml). Protein samples were run on 5 to 8% acryl amide gels and blotted on nitrocellulose. Blots were blocked overnight in 5% milk and incubated overnight with anti-Gnn antibody (diluted 1:1,000). Proteins were detected with horseradish peroxidase-conjugated swine anti-rabbit serum (DAKO, Glostrup, Denmark), followed by an enhanced luminescence reaction.

Immunohistochemistry. Cytospins were fixed with acetone, treated for 10 min with 0.5% hydrogen peroxide in PBS to block endogenous peroxidases, washed in PBS containing 0.05% Tween 20 (PBT), and blocked for 30 min in PBT containing 1% bovine serum albumin. Samples were incubated overnight with anti-Gnn antibody (diluted 1:100) in blocking buffer. Horseradish peroxidase-conjugated swine anti-rabbit antibody was used as the secondary antibody. Samples were developed with diaminobenzidine and enclosed in Entellan (Electron Microscopy Sciences, Hatfield, Pa.).

Quantitative RT-PCR. Quantitative RT-PCR on randomly primed cDNA was performed as previously described (22). The primers used for amplification of *Gnn* sequences were 5′-CAAAGGGCTTCCTGTCAGATT-3′ and 5′-TGCTTG CAGTTCTCCACAAA-3′ for the product of exons 4 and 5 and 5′-CTGAATC CAGACGCCATTTT-3′ and 5′-ATGGCAAAGCTTGGGTCATA-3′ for the product of exons 19 and 20.

Web links for sequence analysis. Sequences were BLAST searched against the mouse and human genome sequence with the Celera Discovery system (Celera Genomics, Rockville, Md.), the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov), and the Public Mouse Genome Sequencing Consortium Database (http://www.ensembl.org). Analysis of protein sequences was done with ScanProsite (http://www.expasy.org/tools/scanprosite) and Pfam (http://www.sanger.ac.uk/Software/Pfam).

Nucleotide sequence accession numbers. *Gnn* sequences have been deposited in GenBank with the following accession numbers: AY651019 (*Gnn* cDNA1), AY651020 (*Gnn* cDNA2), AY651021 (*Gnn* cDNA3), and AY651022 (*Gnn* cDNA4b).

RESULTS

Identification of the novel gene (*Gnn*) near the *Evi12* locus. To begin our search for novel genes near the *Evi12* locus, we applied a previously described exon trapping system (19) to genomic DNA fragments neighboring *Evi12*. This resulted in the identification of an exon (that turned out to be *Gnn* exon 3) located upstream of *Grp94*/*Tra1* (Fig. 1), a gene previously identified near the *Evi12* locus (20). We then performed 3′ and 5′ RACE reactions on an embryonic day 17.5 cDNA library resulting in the isolation of a complete 1,965-bp cDNA sequence (*Gnn* cDNA1) harboring an open reading frame encoding a 409-amino-acid sequence of a protein (Gnn1) with a predicted size of 47 kDa (Fig. 1 and 2).

We were also able to amplify *Gnn* cDNA1 sequences from various other cDNA libraries that we prepared from adult organs and cell lines (for example, NFS107 cells), and we found three alternative versions of the 5′ untranslated region (5′ UTR) that lack exon 2 and/or exon 3. Subsequent 3′ RACE and RT-PCRs performed on a cDNA library made from NFS107 cells revealed the existence of two cDNAs that bear additional 3′ exons. *Gnn* cDNA2 and cDNA3 (Fig. 1 and 2) produce open reading frames encoding 600- and 1,318-amino-acid sequences of proteins with predicted sizes of approximately 69 and 150 kDa (Gnn2 and Gnn3), respectively.

Searches in the NCBI, Ensembl, and Celera mouse databases revealed that all three complete transcripts are located near the *Evi12* locus and confirm the location on mouse chromosome 10C1-2 upstream of *Grp94*/*Tra1* (Celera contig GA_x6K02T2PNY8). *Gnn* cDNA1 to cDNA3 comprise 6, 8, and 18 exons, respectively, and they share the same translation initiation site that is located in exon 4 (Fig. 1 and 2). *Gnn* cDNA1, and most likely also the 5′ end of the other two cDNAs, is similar to the sequence of the Celera gene mCG114872 and to the gene encoding NCBI hypothetical protein MGC40847. Partial sequence similarity was found between cDNA3 and putative Celera genes mCG8343 and mCG8344. The 3′ end of exon 18 of *Gnn* cDNA3 shows 100% sequence similarity to a sequence expressed in *Mus musculus* (accession number AI449705). By searching in the human sequence databases, we found that most parts of cDNA1 to cDNA3 are conserved at least 65% at the nucleotide level in mice and humans.

The cDNA isoforms of the newly identified gene in *Evi12* encode cytoplasmic proteins of which two contain an AAA-ATPase domain. We searched the Pfam (1), ScanProsite (5), and CDART (6) protein homology databases with the novel Gnn sequences. Only one domain, an AAA-ATPase domain encoding amino acids 338 to 543 was found to be present in *Gnn* cDNA2 and cDNA3 (Fig. 2). Furthermore, we identified putative phosphorylation, glycosylation, amidation, and myristylation motifs, and a nuclear targeting sequence in Gnn. Expression constructs encoding fusion proteins between green fluorescent protein (GFP) and the three Gnn isoforms were generated and introduced into HEK-293 cells to study the subcellular localization of the proteins. GFP-Gnn1 to GFP-Gnn3 appeared to be present mainly in the cytosol (Fig. 3A and B). Staining of cytospins of the NFS107 cell line (Fig. 3C) and paraffin sections of adult murine organs (not shown) with Gnn antibody (see below) confirmed the cytosolic localization of Gnn proteins.

Gnn proteins of various sizes are detected in the embryo and in adult mouse tissues. *Gnn* transcripts were detected at very low levels in any organ or cell line by Northern blot analysis using various *Gnn* cDNA probes. However, by applying RNase protection, we could detect relatively high levels of *Gnn* in heart, brain, kidney, and liver, whereas only low levels of expression were observed in spleen, bone marrow, lymph nodes, and thymus (data not shown).

Western blot analysis of lysates from HEK-293 cells transfected with PLNCX-*Gnn* cDNA1 to cDNA3 revealed that a purified polyclonal antibody that was raised against a con-

A

Gnn cDNA3

```
1      cttccggcttctagcgaacgtttagtccgtgactaggggaaacggtagccaagacaatgcgtcctttacccttttctcgtctacgacgcgggtcacgtta
101    cgagtcctttcagcgctacgttcgccgtcacctttcttttgggcgtttcttttcgcaaaagagaaatttgtgaggggttgaatatccggaatgcgtttgg
201    atggtataagcaagtattgtgaaactgtgactgcatcttaactggccagtgggaggtagatcgggctgacctgggaggaactgagaatgagatttgtcct
301    ccaaaataaagctggggctggaatttcactggtgaacaatgaagatatggccctttacccacagaccatgaggaggtaattaatagtggaggcatgcta
401    ttgcttaaccattgcatcccttccacaagaaacaatttaaagaagtgatttgaatgaaactgaccaattgcttatagcaaacacagttaaaagaagaccc
501    aaactgtggaccttggcgactgggccaggatggaacttcctttcttctaagtgcgacatcatcctggagccgagtgaaATGAGCGACGAGGCAAGTGAGA
601    CCGGACAAAGATACAACGGTCAACCCATTTTAAAGCGCCAGAAACCCATCCTACCTTACATATGTTCCACTCTAGATTTTCAAGAAGAGAGAGACTTTTT
701    GGCCAAAAGCATCTTCCCTCGGCTTAATGACATCTGCAGCTCCCGGGGCACCTACTTCAAAGCTGTGGACTTGAGATGGTCAGCTGTGAAGGCCCATAAG
801    TCCTTCACCAGCAACCAGTTCCGACAGTACTCCTGTCTCCAGTCTCAACACCTGAAACTCTCCCTGGACTACGTGAACAGATGCTTCCCGTTTTTCATCG
901    GCCTGCTGGGGCAGACCTACGGAGATTTCCTCCCCGACTACACACCTTTCCTGTTATCCCAAGTGAAGGACTTCGAAAGTTTATCCAAGGGAAAAAAGAA
1001   TCTATACATTGCTGCCAAAAATGGTTACCCTTGGGTTCTCAAGACTCCCAACTGCAGCCTGACAGAGTTCGAGATCATCCAAGCAGTATTCCGGAAGAAA
1101   TCTCAATTTCAATTTTTCTACTTCCGGACATCAAATTCGCTGCTGCGAACTTTTAATGAGGAAGAGGAGGAGGAAGAGGAGAAGCTGTCCTCAGCATATC
1201   TGTTGAACGAACAGGGGAAGATGAAGGTTGGAAAGCTCAAGGCTAAGATCATTGGCAAAGGGCTTCCTGTCAGATTCTACAGAGACCTGGAGGAACTGGG
1301   GGATATGGTTTGGAAGGACTGGTCGGCTGTTGTGGAAAAGCTCTATCCATTCACTACGATCATGGGAAATATAGACTACAAACACAGTTTCGAGAATTTG
1401   TATCATGAAGAGTTTGTGGAGAACTGCAAGCAAGTTTTTGTTACTTCCAAGGAGTCAAACAGAACCTTTGAAATATTGGAAAGATTTGCTATAAAAGATC
1501   TTGATCTTGATCTTGATACTGACAGTACTATAGCAGGCTCGGGGTTAGACTCCATTCTCAGAATCAATTCCCTTCCAACTTGTAAGTCCATTTTGCTCTT
1601   GTCTGGGGAACGCGGCTGTGGAAGTCCACACTGATTGCCAACTGGGTCAGTAATTTCCAAAGCAAACACCCCGGAGTGCTGATGATCCCATACTTTGTG
1701   GGCAGTACGTGTGAGAGCTGTGACATCATGTCTGTGATCCACTACTTCGTCATGGAGCTTCAGCACAGAGCCAACGGTCCCCGGCTTGAAATGGATTTCC
1801   TTAACGAGGACTCAAATGTCTTGGTCTTTTCACTTCTCGTAGAAGTGTTCATAGCCGCCATAAGCTTAAAGCCATGCATCCTGGTACTAGACGGGATCGA
1901   AGAGCTGATCGGTATCTATGGGATTTCAGGTCAGAAGGCGAAAGATTTCTCCTGGCTGCCGCGCTCCCTGCCTCCTCACTGTAAATTCATTCTGAGCAGC
2001   GTCTCCTCCAGTCTGTCCTGCAAGTCGCTGTGTGCCCGCCCTGACGTGAAGATCGTGGAACTCAACAGCATCGGGGACGAAGACACCAAGTTCAACATCT
2101   TCAGACAGCACCTCTCCCCTGCCGACCAAGAACGCTTCGGGCAGAGCAAGCCCATTTTGAGGAAGAAACCAAACCTGAGCCCTCTGAAGCTCGCAATCAT
2201   CGCCAGCGAGCTGCAGGAGTGCAAAATCTACCGCAATGAGTTCCAGTGTCTCCGGGAGTACTTGGAGGTTGCCTCTGTGCAGGAGCTCTGGGAGTTGATT
2301   CTGAAGCGCTGGGTTGAAGATTATAGTTGGACTTTGAAGCCTAAAGACACAACTCTAGACACCGTGATTCCAGGGCCAAGTGGCTGGGTAGTGGATGTGC
2401   TGTGCTTGCTCTGCATCTCTCACTGCGGGCTGGCTGAGGATGAACTGCTCCAGCTTTTGGACACGATGGGCTACAGGGACCACCACAAAGTGACGGCGGT
2501   GCACTGGGCAGCCTTCCGCCAAGCCACCAAAACCTGGATCCAGGAGAAGCCCAATGGTCTCCTCTACTTCCAGCACCAGTCCCTAAGGAGTGCCGTGGAG
2601   CACAAGCTGCTGGGTGTAAGCACTCCGGTGAGAGAGAGCAATCCCAATGTGGCCCAGAACAGCGTGAATCACAAGAAGGCACATTTCCACCAGGTCTTGA
2701   TGAGGTTCTTCCAGCGGCAAACCATCTTCTGGAGGGTGTATCAGGAGCTGCCCTGGCACATGAAGATGAGCGGGTATTGGGAAGGTCTATGTAACTTCAT
2801   CACGAACCCCAGCATCACAGATTTCATATCGAAAATCCAAAACCCAAGCTTGTGGACCAGGCTGCACCTTGTCCACTACTGGGATGTGCTGCTGGAAGCC
2901   GGCAATGACGTGTCTGAGGCTTTTCTGCTCTCTGTTGCCAAGATAGAAGGGGAACAATTCCAGAAACTCAAGAAGCGAACCACACTCTCAGTGCTGGAAT
3001   GCAGCCTGTCCGAGATTACTGCTGCTGATAAAGGCAGAATTATCCTCTTTATTGGAAGTTTCCTGAAGCTAATGGGCAAGATCAATGAAGCTGAAAAGCT
3101   GTTCTTGAGCGCTGAGGACTTGTTACTACAGAGCCCATCCATGACAGAAATGCTGCTCAGAGCTCAGAATGCCATTGGGGAATTATATCTTGAGATTGGG
3201   ATGACTCCGAAAGGACTCACATATTTTCAGAAAGCTTGGTCAAATCTGCTGCGGTTTACACTCAGTGACCTAAAGATCAGCCAGGAATTGATGAAGCAGA
3301   AAGTTAAAGTGATGAATAACCTGGCAGAATCGGCGCCTGGGGAATTCTTAAAAGAAAACCACGTTCTGGAATATGCTACCGAAATCTCCAAGTACGTGAC
3401   TGGTAATCCCCGTGATCATGCTACCATGAAATATACTGAAGGTGTTCTCATGTTGGCTTCCGGAAACGCAGCCCTGGCAAAACTCAAGTTTCAAGAGTGT
3501   TTAACTATCAGAAGATGGCTATTTGGCAATAAAAACATACTAGTTGGAGAAATTATGGAATTCTTAGCAGATTTACTATTTTTTCTACTAGGAGAAAATG
3601   AAAAATCTCAAAAGAAGCAAGCAATTGAATATTATAAACAAGTCATAAAAATCAAGGAGAAGGCAGATACGGTGGCCACCTGCAAGCTTGTGAGGAAGCA
3701   TCTGAGTATAAGCCTCAGCGACACCTTGTGTAAACTAGCAGGCCAGCTGTTGTCAGGTGACTTCTGCCATCATGCCACAATGGAGGCAGTCAGCTATTTG
3801   TATAGGTCACTTGATTTAAGGGCAGCTCACCTGGGGCCTACCCACGCTTCCATTGAGGGAATACTACACCTTTTACGGGAAATCCAGAGGTCCCGGGGCA
3901   GGAGGTCTTGGCCTCAAAGCATGAACCATCTATTCCCTAACGGTTCTAGGAATGGCTTTTCATTATGGGAGAATGTGCCTAAATTAAACTTCCACAGCGC
4001   TCAGAGTTCTGACACGGTAAACACTGCAATGTGTATGAATATACGTAGGTTTCAGAGAGTTAAAAGCACACAGCCTTCTCTGGTTTCAGATAAACCAAAA
4101   TATGTTCCCGGCAAAGGAAAGAAGACCTTGGCTCCAATTCTGTGCAAGTCTGCTGAGGAAAAGTTCCAACGTCAGGCTTCAGACTCACAAATATGGAATA
4201   GTCCAAGAAGACAACCTGCCAGGAAAAAGGCAGCCTGTCCCCTTAAGACGGTCTCGCTCATTGACAAGAACGGCTTGGTGAGACTCTCAAGGCAGAGTGT
4301   TTCTTCTGCTGAGCTGGACAGCAGAAAGGGCCTGATCACTTCCATCTGCCGGCAACCCCTGCAGCGACCTCATAATGTGGACAATCCCTGGAAGTCCATA
4401   TCGGAACTCGTGTCAGAAAAGTGGCTCTTCCACACTCCTCAGTACTGCTTCACTCCTCAGAAGCCAGGCTTCCCAAGAAGATCTCAGATTGAATCAAAAT
4501   TGCTGAAGACCTCAGATGACCCCAATAAGGAATAAaaataattttgggctgttttgt
```

B

remaining part *Gnn* cDNA1

```
1779   AACCGGGCCAGCCTTATGGGCAGAAAGTGAagacaagtcccctgctagactctctgaaccccactttactttatataaataccaacaagactgctttaa
1879   acacctcccatctcgcttgtcaaagtaaagccagaaagttcctaagaccagaaaacaactggttttgattaaaaaaaaaaaaaaaa
```

C

remaining part *Gnn* cDNA2

```
2373   GGTGTGTGAtagctgatcctaaaagtgcctcctttatggctccaagacttgtttgtcccagtttctcggttgttctgtttctgctgccccctgctggtgc
2473   agtccctcatcctcaccaaggacaatcttgttttgtttccacagggtctctctatgtagttctggctagaacttgcttatgtagaccaggctggactcaa
2573   actcacacagctccatctgcttctgcctcccgagttctgggactaaaagggtgtgctatataccttggctctcctggcaccttgatgcctgatttcctct
2673   tcaactttctcctagacctttctcagtttagtaacagctctttccttcctagttggggtaaaacaaacaaacaaacaaacaaacaaaaaaaaaaaaaa
```

D

Gnn cDNA4b (partial cDNA)

```
1      CCTCAGCGACACCTTGTGTAAACTAGAATTGGTTCCATCCATCATGAACAACTTGCTGAATCCAGACGCCATTTTCTCGAACAATGAGATGAGCCTGTCA
101    GACATCGAAATATATGGCTTTGATTATGATTACACCCTGGTATTCTATTCCAAGCACCTCCACACACTCATCTTCAATGCTGCCCGGGACCTTCTCATCA
201    ACGAACACAGGTATCCTGTGGAAATCAGGAAGTATGAGTATGACCCAAGCTTTGCCATCCGGGGACTCCATTACGACGTCCAGCGGGCAGTGTTGATGAA
301    GATCGATGCTTTTCATTACATCCAGATGGGGACAGTGTACAAAGGCCTCAGTGTTGTCCCTGATGAAGAAGTCATAGACATGTATGAGGGGTCCCACGTG
401    CCCTTGGAGCAGATGAGTGACTTCTATGGAAAGAGTTCTCATGGGAACACCATGGAGCAGTTCATGGATATCTTCTCGCTGCCCGAGATGACCCTGCTGT
501    CCTGTGTGAACGAGCACTTCCTGAAGAACAACATTGACTATGAACCTGTGCACCTGTACAAAGACGTCAAGGACTCCATAAGGGATGTCCACATCAAAGG
601    GATAATGTACCGAGCAATCGAAGCGGATATTGAAAAGTACATCTGTTATGCTGATCAGACCCGTGCAGTGTCGGCTAAACTGGCTGCCCACGGCAAGAAG
701    ATGTTCCTCATTACCAATAGCCCAAGTAGCTTTGTAGACAAAGGGATGCGGTACATCGTGGGGAAGGACTGGCGAGACCTGTTCGATGTGGTCATAGTTC
801    AGGCCGAGAAACCCAACTTCTTCAATGACAAGCGGAGGCCTTTCCGAAAGGTGAATGAGAAAGGTGTCCTACTCTGGGATAAAATCCACAAGCTGCAGAA
901    AGGCCAGATTTACAAGCAGGGCAATTTGTATGAATTTTTGAAGCTCACTGGATGGAGAGGATCCAAAGTACTGTATTTTGGTGACCACATCTACAGCGAT
1001   CTGGCGGATCTGACCCTCAAGCACGGCCGGCGCACTGGAGCGATCATCCCCGAGCTCCGTTCCGAGCTCAGGATCATGAACACGGAGCAGTACATCCAGA
1101   CCATGACCCGGCTGCAGACCTTGACGGGGCTCCTGGAGCAGATGCAGGTCCACAGAGACGCCGAGTCACAGCTGGTTTTACAGGAGTGGAAAAAGGAGAG
1201   GAAGGAGATGAGAGAAATGACCAAAAGTTTCTTCAACGCCCAGTTTGGGAGCTT
```

FIG. 2. (A) Nucleic acid sequence of *Gnn* cDNA3. Translated (uppercase type) and nontranslated (lowercase) sequences and start and stop codons (bold type) are indicated. The AAA-ATPase domain that is present in cDNA2 and cDNA3 is underlined. *Gnn* cDNA1 and cDNA2 are identical to cDNA3 up to positions 1778 and 2372, respectively (the last three identical nucleotides are indicated by the grey shading). (B and C) Remaining 3′ sequences of cDNA1 and cDNA2 (representing longer versions of exon 6 and exon 8). (D) Sequence of *Gnn* cDNA4b. The last three nucleotides before the fusion between exon 16 and *TU12B1-TY* as found in *Gnn* cDNA4a or cDNA4b are indicated by the grey shading (see also the box at positions 3735 to 3737 in panel A).

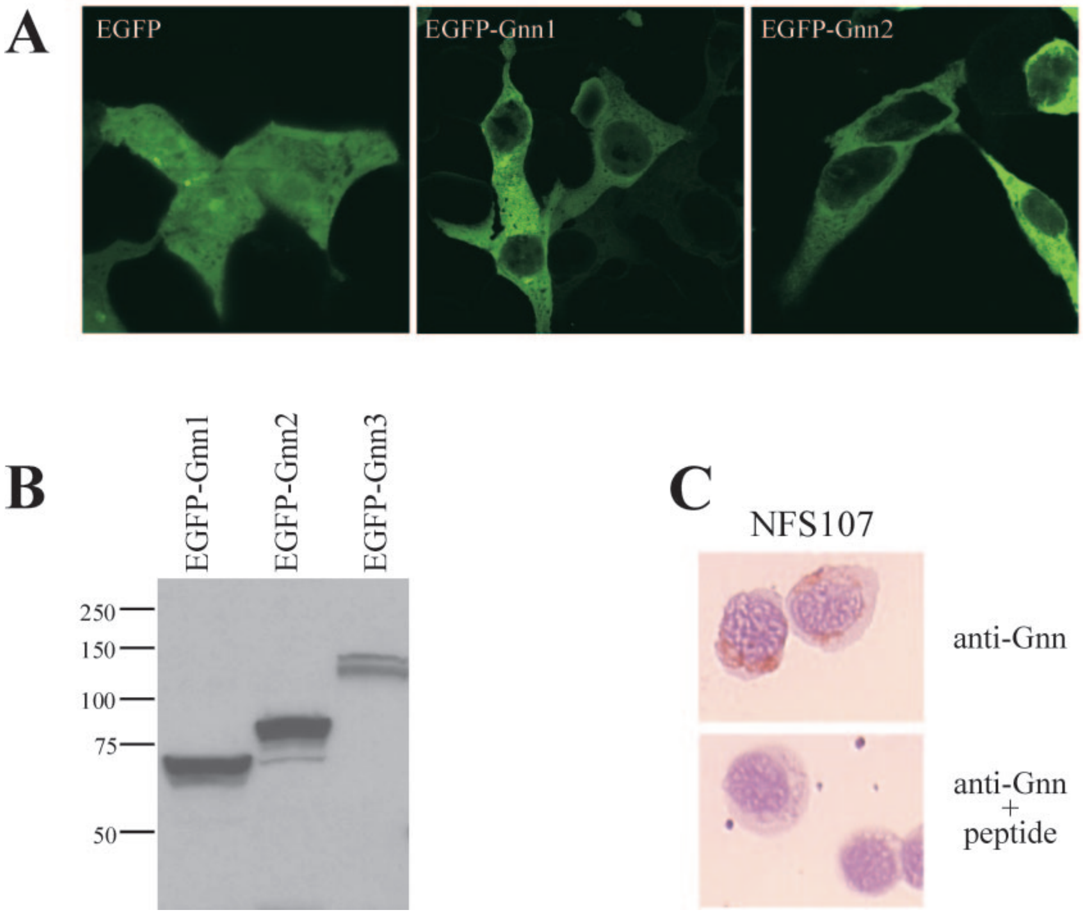


FIG. 3. (A) Subcellular localization of Gnn. HEK-293 cells were transfected with constructs encoding EGFP, EGFP-*Gnn* cDNA1 (EGFP-Gnn1), EGFP-*Gnn* cDNA2 (EGFP-Gnn2), and EGFP-*Gnn* cDNA3 (EGFP-Gnn3). While EGFP is expressed within the entire cell, EGFP-Gnn fusion proteins localize to the cytoplasm as shown for EGFP-Gnn1 and EGFP-Gnn2. (B) Western blot analysis with GFP antibody of lysates of 0.5×10^6 HEK-293 cells transfected with EGFP-Gnn constructs demonstrating the three different fusion proteins, EGFP-Gnn1 to EGFP-Gnn3. The positions of molecular mass markers (in kilodaltons) are shown to the left of the blot. (C) Immunostaining of NFS107 cells with Gnn antibody confirmed the cytosolic localization of Gnn protein. Specificity is shown by a control staining using Gnn antibody preincubated with an excess of the peptide against which it was raised, resulting in total loss of the signal.

served N-terminal peptide specifically recognized the Gnn1 to Gnn3 protein variants (Fig. 4). Although the Gnn3 protein has a predicted size of 150 kDa, it clearly ran faster on Western blots. This may be due to the conformation or to posttranslational modification of the protein.

Western blot analysis revealed Gnn protein expression in most analyzed tissues. Proteins of various sizes ranging from ~25 to ~250 kDa could be detected in adult organs and embryos, whereas almost no Gnn protein was detected in bone marrow and spleen (Fig. 5). The largest Gnn proteins were present in day-9.5 and day-11.5 embryos. A slightly smaller protein was present at low levels in brain and muscle, while a ~150-kDa band was highly expressed in lung. Expression of Gnn proteins of ~65 and ~55 kDa was evident in most organs with particularly high expression of the ~65-kDa protein in heart, kidney, thymus, stomach, and liver. Multiple other bands of variable molecular sizes were observed in most tissues that we investigated. All these variants could represent (modified) Gnn splice variants or Gnn breakdown products. We compared the sizes of the Gnn1 to Gnn3 proteins and those detected in the murine organs. The Gnn1 and Gnn2 proteins, respectively, have similar sizes as a ~45-kDa protein that was detected in heart and a ~65-kDa protein that was detected in various tissues, suggesting that they represent identical proteins. The Gnn3 protein, however, migrated faster than the protein found in lung (and in NFS107 cells [see below]) that migrates at the 150-kDa position (data not shown).

A Gnn protein variant is highly expressed in the *Evi12*-containing cell line NFS107. Western blot analysis was performed on protein extracts from various murine myeloid leukemia cell lines. A strikingly strong signal representing a protein of approximately 150 kDa (Gnn^{107}) was identified in NFS107, the myeloid leukemia cell line harboring a proviral insertion in *Evi12* (Fig. 5 and 6). The Gnn^{107} protein migrated slower than the Gnn3 protein, which has an expected size of 150 kDa (data not shown), suggesting that it is a larger protein. Increased blotting periods or running the samples on a low percentage (5%) polyacrylamide gel, which both allow improved blotting of large-molecular-size proteins, resulted in detection of two additional large proteins of approximately 225

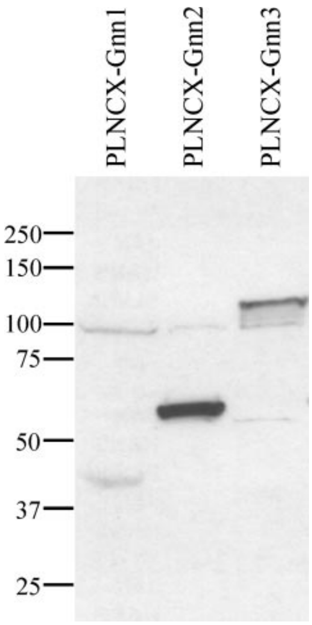


FIG. 4. An antibody raised against the N terminus of Gnn specifically recognizes the untagged Gnn1 to Gnn3 isoforms (with predicted sizes of 47, 69, and 150 kDa, respectively). The Gnn3 protein clearly migrates faster than expected. For this experiment, lysates of 0.5 × 10^6 HEK-293 cells transfected with constructs PLNCX-Gnn1 to PLNCX-Gnn3 were used. The positions of molecular mass markers (in kilodaltons) are shown to the left of the blot.

and 300 kDa in NFS107 cells (Fig. 6B). Interestingly, NFS36 cells also exhibited expression of two high-molecular-weight proteins, one with a size similar to that of the 225-kDa protein expressed in NFS107 cells. Binding of Gnn antibody to most bands appeared to be specific, since most proteins were not visible after preincubation of the Gnn antibody with excess Gnn peptide. The exception was a 90-kDa protein that was detected in some of the experiments (Fig. 6B and C). The fact that expression of the Gnn^{107} isoform was not observed in various other NFS cell lines (Fig. 6) or in bone marrow or spleen (Fig. 5) suggests that the high level of expression of this isoform in NFS107 cells may be the result of the proviral insertion near *Gnn* in *Evi12*. We also performed Western blot analysis on a number of primary tumors containing a VIS in *Evi12*, including two AKXD tumors in which the VIS has the same orientation as the one found in NFS107 cells (17). We were not able to detect aberrant Gnn protein expression, e.g., expression of the Gnn^{107} isoform in these tumors (data not shown), which may be due to the fact that these tumors are oligoclonal.

The putative 5′-nucleotidase-encoding gene *TU12B1-TY* is part of the novel *Gnn* gene. Since Western blot analysis revealed that proteins larger than the Gnn3 protein encoded by *Gnn* cDNA3 were present in NFS107 and NFS36 cells and in mouse embryonic tissues, we searched for additional *Gnn* sequences in several available databases. The direct Celera neighbor of the mCG8344 gene is a putative gene designated mCG8347 (Fig. 1), encoding the mouse homolog of the human TU12B1-TY protein, a protein related to the 5′-nucleotidase family of enzymes.

By using RT-PCR and primers for *Gnn* exon 6 and the first exon of *TU12B1-TY* (exon 19), we were able to amplify a specific in-frame fusion product between the two genes designated *Gnn* cDNA4a (Fig. 1). The fact that this product could be amplified only from the cDNA of NFS107 cells, and not from cDNA of control cell lines (Fig. 7), strongly suggests that *TU12B1-TY* sequences are part of the Gnn^{107} protein. From the cDNA of NFS107 cells, we also amplified the *Gnn* cDNA4b product (Fig. 1 and 2), using primers for exon 16 of

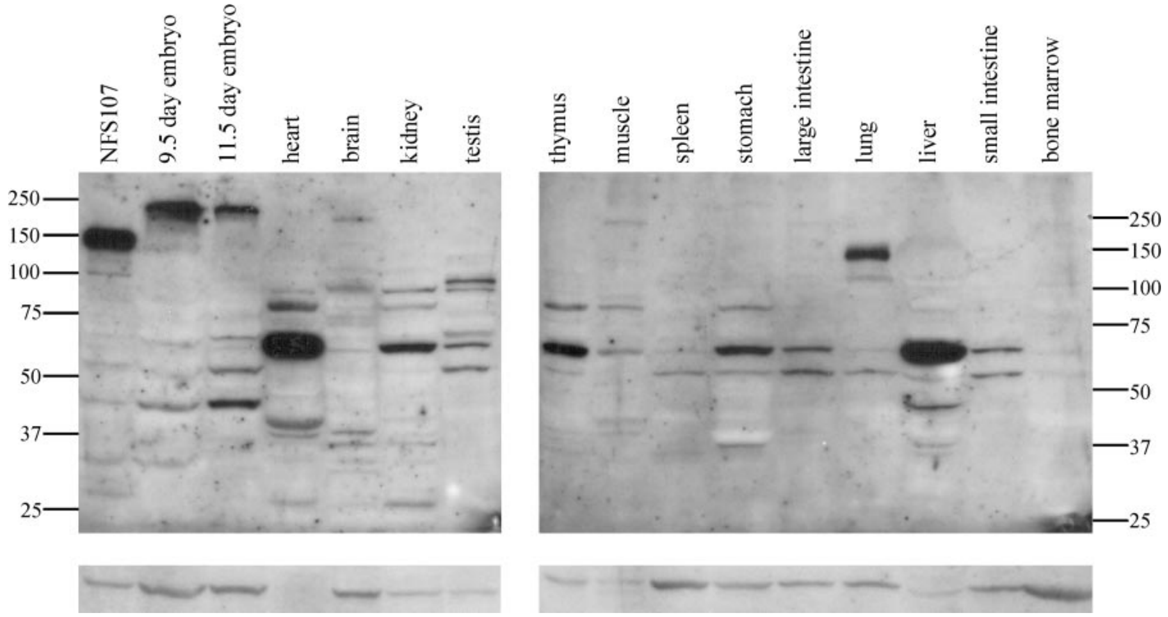


FIG. 5. Western blot showing expression of multiple Gnn isoforms in NFS107 cells, mouse embryos (embryonic days 9.5 and 11.5), and in various adult mouse tissues. The ~150-kDa protein (Gnn^{107}) that is detected in NFS107 cells is absent in bone marrow and spleen. The positions of molecular mass markers (in kilodaltons) are shown at the sides of the blot. In the bottom blots, actin was used as a control to ensure equal loads.

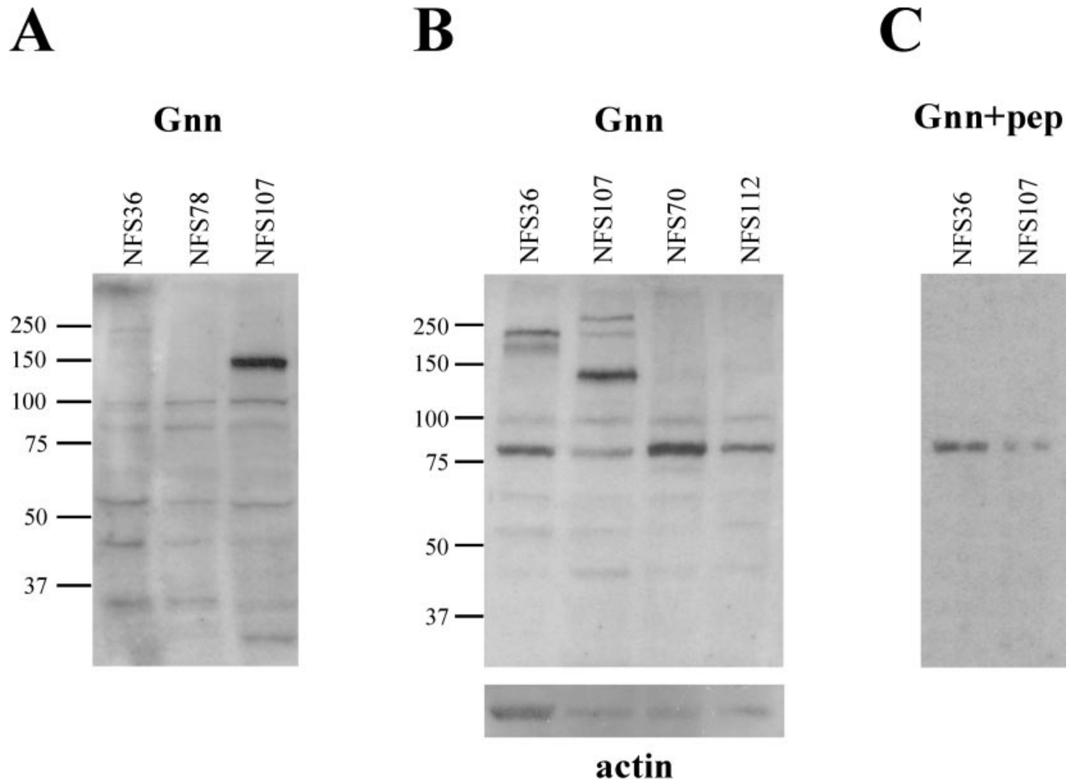


FIG. 6. A 150-kDa protein (Gnn^{107}) specifically recognized by Gnn antibody is overexpressed in NFS107 cells compared to control cell lines. Two independent experiments are shown (A and B). Clear overexpression of Gnn^{107} is visible in both experiments. Two additional bands are visible in panel B in the NFS107 lane, because longer blotting times resulted in enhanced transfer of very large proteins. Moreover, NFS36 cells also show expression of two high-molecular-weight bands. The blot in panel B was stripped and incubated with anti-actin antibody as a control to ensure equal loads. The positions of molecular mass markers (in kilodaltons) are shown to the left of the blots. (C) Experiment performed in parallel with the experiment in panel B showing that preincubation of Gnn antibody with excess blocking peptide (Gnn+pep) results in disappearance of specific bands in NFS36 and NFS107 cells. A ~90-kDa protein background band remains visible in this experiment.

cDNA3 and the last coding exon of mCG8347. The sequence of this product overlaps with the sequence of the *Gnn* cDNA4a product (both contain an in-frame fusion of exon 16 with the first exon of mCG8347), adding 13 additional exons (exons 19 to 31) to the *Gnn* sequence. We also identified a number of alternative cDNA products with this PCR which lack different mCG8347 exons (data not shown). We have not been able to amplify a full-length *Gnn* cDNA4 product using primers for exon 4 and exon 31, possibly due to the restricted sizes of the cDNA fragments in the NFS107 library. The largest possible full-length *Gnn* product, i.e., exons 4 to 31 (excluding exons 17 and 18), that includes TU12B1-TY sequences would be predicted to encode a protein with an estimated size of ~180 kDa. Since Gnn proteins migrate a bit faster than expected, such a protein could represent the Gnn^{107} protein.

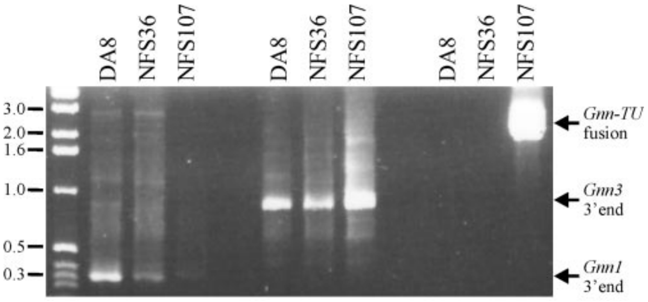


FIG. 7. A fusion cDNA containing *Gnn* and *TU12B1-TY* sequences is specifically expressed in NFS107 cells. RT-PCR was performed with three different primer sets on cDNA from NFS107 and control cell lines. While there is no increase in the amplification of the 3′ end of *Gnn* cDNA1 (left) and cDNA3 (middle) in NFS107 cells compared to the other cell lines, the *Gnn*/*TU12B1-TY* fusion (*Gnn-TU*) cDNA4a (right) was amplified only from NFS107 cells. The positions of molecular size markers (in kilobases) are shown to the left of the blot.

We performed quantitative RT-PCR with primer sets in exons 4 and 5 and exons 19 and 20 of *Gnn*/*TU12B1-TY* (Table 1). NFS107 cells had the highest levels of expression of exons 4 and 5, which may be part of *Gnn* cDNA1 to cDNA4 and possibly other *Gnn* variants containing the N terminus. Exons 19 and 20 may be part of cDNA products that contain *TU12B1-TY* sequences, either with or without upstream *Gnn* exons. NFS107 cells did not have the highest levels of expression of exons 19 and 20 (Table 1). This suggests that the VIS in *Evi12* does not enhance the total expression of *TU12B1-TY* mRNA, e.g., via activation of an internal promoter.

***GNN* is overexpressed in human AML.** To determine the expression of *GNN* in primary human AML, we screened the database containing the genetic profiles of 285 cases of primary AML, which were recently generated using the human U133A Gene Chip from Affymetrix (21). Although probe sets corre-

TABLE 1. Relative levels of products of exons 4 and 5 and exons 19 and 20 in NFS107 and control tumor cell lines determined by quantitative RT-PCR

Cell line	Exons 4 and 5			Exons 19 and 20			C_T RI[d]
	C_T[a]	ΔC_T[b]	Expression relative to NFS107[c]	C_T	ΔC_T	Expression relative to NFS107	
NFS107	26.26	0.81	1.00	27.77	2.32	1.00	25.45
NFS36	28.20	2.85	0.24	28.76	3.41	0.47	25.35
NFS56	27.68	2.2	0.38	26.73	1.25	2.10	25.48
NFS124	27.16	1.04	0.85	27.61	1.49	1.78	26.12
DA8	30.64	1.65	0.56	30.77	1.78	1.45	28.99

[a] C_T, threshold cycles (the number of PCR cycles to reach the threshold level).

[b] ΔC_T, C_T value after normalization with endogenous reference = C_T of the product of interest − C_T of RNase inhibitor.

[c] Calculated by using the formula $2^{\Delta C_T(\text{NFS107}) - \Delta C_T\text{ (cell line)}}$.

[d] RNase inhibitor (RI) was used as the endogenous reference.

sponding to the human counterparts of the MGC40847 and AI449705 sequences were absent on this array, a probe set unique for a piece of the very long 3′ UTR of the human *TU12B1-TY* gene was present. We compared the levels of expression of *TU12B1-TY* in 16 predefined AML clusters (21). Strikingly, two clusters showed increased *TU12B1-TY* expression compared to all other clusters (Fig. 8). To demonstrate that the expression in these two clusters was significantly different from expression in the rest of the AML patients, we performed significance analysis of microarrays (18). This analysis indicated that in cluster 6, the average expression is increased 2.4-fold compared to the geometric mean expression of all AML patients (score [the absolute difference between the average for the cluster and the average for all 285 AML samples, divided by the standard deviation of the differences between the averages for the two groups in 300 permutations] of 6.2; *q* value [the chance that the observed change is caused by coincidence] of 4.7%), while in cluster 13, a 2.1-fold increase is found (score of 9.5; *q* value of 0.08%). Thus, the expression of *TU12B1-TY* is indeed significantly increased in these two clusters. One of these two clusters, cluster 13, contains all patients with a t(8;21) translocation. The other group of patients expressing high levels of *TU12B1-TY* includes cluster 6 and part of cluster 7 (21). The subset of patients included in cluster 7 appeared to have common molecular signatures in previous analyses (21), demonstrating that these patients may be similar to the patients in cluster 6. Interestingly, all these patients, including those in cluster 7, have a normal karyotype and most have a *FLT3* internal tandem duplication (21). Quantitative RT-PCR on human cDNA samples using a primer set located in the coding region of human *TU12B1-TY* confirmed the increased expression of *GNN/TU12B1-TY* in patients in clusters 6, 7, and 13 (data not shown).

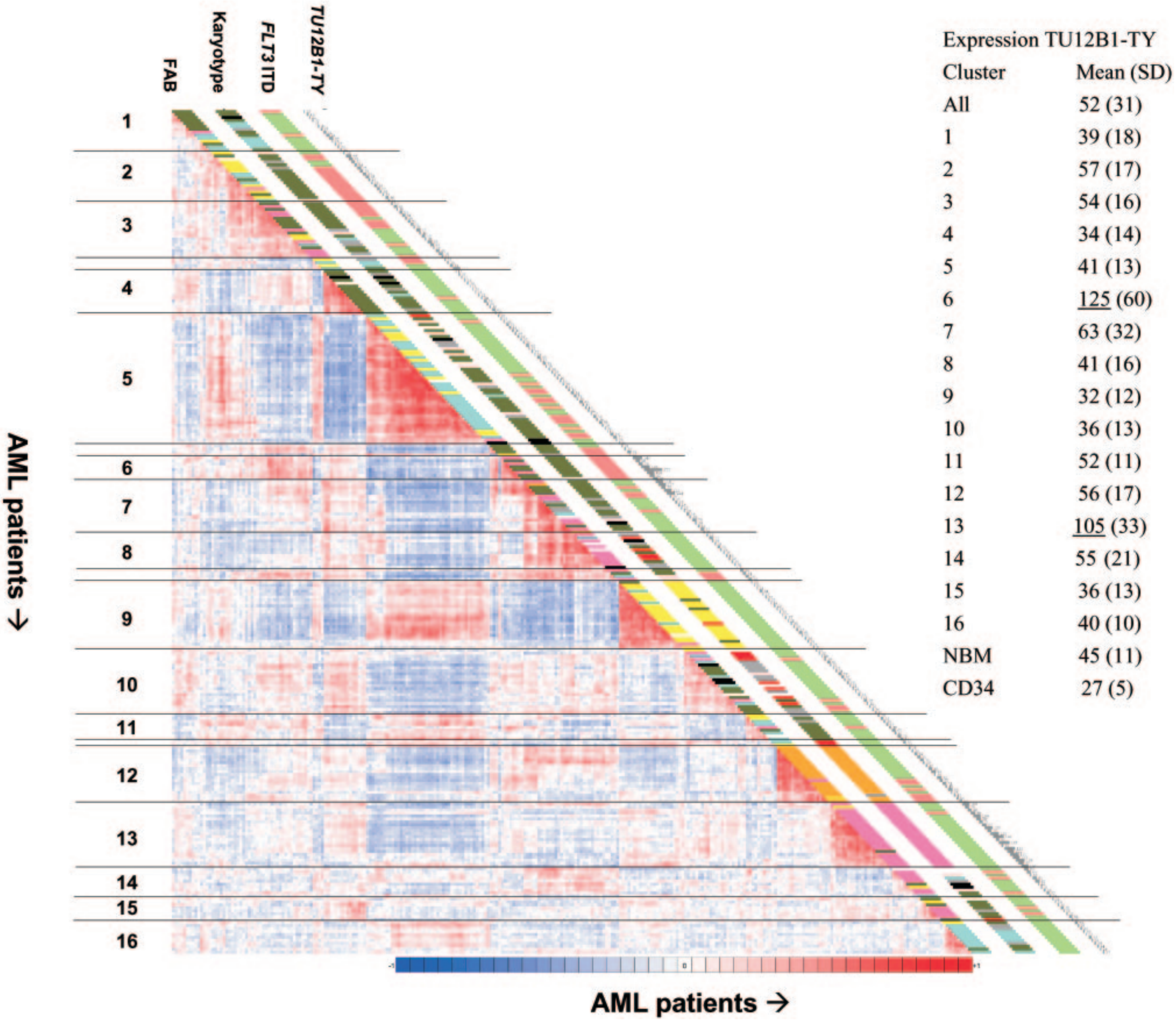
DISCUSSION

We identified *Gnn*, a complex novel gene on mouse chromosome 10, which is conserved in mice and humans. Our data show that the murine *Gnn* gene consists of at least 31 exons and that several isoforms are expressed in murine embryonic and adult tissues and in murine leukemic cell lines. We identified the full-length coding regions of three different *Gnn* cDNAs (cDNA1 to cDNA3) that encode proteins with a predicted size of 47 kDa (Gnn1), 69 kDa (Gnn2), and 150 kDa (Gnn3). We also identified partial cDNAs (*Gnn* cDNA4a and cDNA4b) that contain additional 3′ sequences not present in cDNA1 to cDNA3. Multiple protein isoforms with sizes ranging from approximately 25 to 250 kDa were detected by Western blot analysis with a Gnn-specific N-terminal antibody in embryos and adult organs. According to our data, Gnn is a cytosolic protein that shows tissue-specific expression of the various protein isoforms, with high levels of expression in muscle, liver, and lung. In contrast to NFS107, a leukemia cell line that contains a VIS in *Evi12*, only limited amounts of Gnn protein could be demonstrated in hematopoietic organs, such as bone marrow and spleen.

***Gnn* and *Evi12* in mouse leukemia.** The VIS *Evi12* may have an important role in virally induced murine hematopoietic disorders, since it has been identified as a cVIS in screens using murine leukemia (20) as well as lymphoma models (17). The first three noncoding exons of the newly identified *Gnn* gene encompass the *Evi12* locus, making *Gnn* a likely target for deregulation of expression by retroviral insertion. This notion is supported by the fact that NFS107, the only leukemia cell line containing a viral integration in *Evi12*, is the only tumor in which we observed high levels of expression of a *Gnn/TU12B1-TY* fusion cDNA and of a ~150-kDa protein (the Gnn^{107} isoform) that may represent a Gnn/TU12B1-TY fusion protein. We have not been able to detect this Gnn isoform in primary tumors containing *Evi12*. This may well be due to the fact that these tumors are oligoclonal. The results of several studies indicate that tumors induced by murine leukemia virus contain primary hits, which are involved in tumor initiation and which are present in all tumor cells, while secondary and tertiary hits, involved in tumor progression, are present in only a fraction of the tumor cells (3, 9, 10, 14, 17). In the case of secondary tertiary and hits, generation of in vitro-selected clonal cell lines from freshly isolated primary tumors, such as the NFS107 cell line, is necessary to show changes in expression. Therefore, we propose that *Evi12* represents a secondary or tertiary hit and that *Gnn* is involved in tumor progression.

The retroviral integration in *Evi12* in NFS107 cells is in the reverse orientation to that of transcription of *Gnn*. Such integrations may have an enhancing effect on the transcription of the target gene (for a review, see reference 7). The reverse retroviral integration may also disrupt the normal regulation of *Gnn* translation by the 5′ UTR. It is currently unknown how this could lead to the specific overexpression of the Gnn^{107} isoform. Future studies will focus on the identification of the sequence of the Gnn^{107} protein. This will require generation of additional Gnn antibodies, since the N-terminal antibody we used in this study appeared to be unsuitable for immunoprecipitation. Functional assays using in vitro and in vivo models will be necessary to address the functional significance of *Gnn* in the development of leukemia.

***Evi12* and *Grp94*.** The VIS *Evi12* is located upstream of and near the promoter of the *Grp94/Tra1* gene. Our previous results suggested that *Grp94/Tra1* is not the *Evi12* target gene, since we did not observe altered expression levels of *Grp94/Tra1* in NFS107 cells and various primary tumors containing this virus integration (20). Another mechanism of transformation by proviral insertion is the generation of aberrant fusion



Expression TU12B1-TY

Cluster	Mean (SD)
All	52 (31)
1	39 (18)
2	57 (17)
3	54 (16)
4	34 (14)
5	41 (13)
6	125 (60)
7	63 (32)
8	41 (16)
9	32 (12)
10	36 (13)
11	52 (11)
12	56 (17)
13	105 (33)
14	55 (21)
15	36 (13)
16	40 (10)
NBM	45 (11)
CD34	27 (5)

FIG. 8. Adapted correlation view of data from 285 AML patients. The correlation visualization tool of Omniviz displays pairwise correlations between the samples. The cells in the visualization are colored by Pearson's correlation coefficient values with deeper colors indicating higher positive (red) or negative (blue) correlations. The color scale bar at the bottom of the figure indicates 100 percent positive correlation (red) toward 100 percent negative correlation (blue). The 16 clusters of AML patients, identified on the basis of the correlation view, are indicated by the numbers 1 to 16 to the left of the figure. French-American-British (FAB) classification, karyotype, *FLT3* internal tandem duplication (ITD), and *TU12B1-TY* expression are shown in the columns along the original diagonal of the correlation view. FAB classification is indicated as follows: FAB M0 (red), M1 (green), M2 (purple), M3 (orange), M4 (yellow), M5 (blue), and M6 (grey). The karyotype based on cytogenetics is indicated as follows: normal (green), inv(16) (yellow), t(8;21) (purple), t(15;17) (orange), 11q23 abnormalities (blue), 7(q) abnormalities (red), +8 (pink), complex (black), and other (grey). *FLT3* internal tandem duplication (ITD) mutations are depicted in the same set of columns (positive [red] and negative [green]). The mean expression levels of *TU12B1-TY* (raw Affymetrix U133A GeneChip data) for each cluster with standard deviations (SD) are indicated to the right. The values for clusters 6 and 13 are underlined because they exhibited increased *TU12B1-TY* expression and were significantly different from the values for the other clusters. NBM, normal bone marrow; CD34, CD34$^+$ cells.

transcripts between viral and target gene sequences (7). Although no such fusion or read-through transcripts were apparent by Northern blot analysis (20), we are currently investigating whether such messages may exist in NFS107 cells and other tumors, using more sensitive strategies. If this were the case, then *Evi12* would be an example of a VIS causing the deregulation of two genes that lie in its proximity, i.e., *Gnn* and *Grp94*.

Possible function of Gnn proteins. The only known functional domain that is present in Gnn proteins (i.e., Gnn2 and Gnn3) is an AAA-ATPase domain. AAA family proteins are chaperones that perform diverse functions but have a similar mechanism of action, namely, the energy-dependent unfolding of proteins (for a review, see reference 12). More clues about the function of *Gnn* may come from the partial cDNAs (cDNA4a or cDNA4b) that contain sequences of both *Gnn* cDNA3 and the putative gene encoding the TU12B1-TY protein. This protein shows homology to the cytosolic purine 5′-nucleotidase class of enzymes (15). Nucleotidase enzymes, some of which are strongly regulated by ATP, hydrolyze 5′-(deoxy)ribonucleotides to nucleosides and phosphates. Together with the diphosphate kinase enzymes that are known targets of leukemia virus (i.e., Nm-23 [9; also unpublished data]), they are involved in the maintenance of a constant intracellular composition of nucleotides (for a review, see reference 2). On the basis of our results and the above, we propose that the longest form of *Gnn*/*TU12B1-TY* probably encodes an ATP-regulated cytosolic 5′-nucleotidase.

Our data also suggest that we cloned only four of several *Gnn*/*TU12B1-TY* cDNAs. Additional studies are necessary to address to what extent the proteins encoded by the *Gnn* cDNAs are similar to the proteins detected by Western blot analysis.

***GNN* and human leukemia.** Nucleotidase enzymes are known for their influence on the pharmacological efficacy of antiviral and antitumor nucleoside analogs. This may explain why high levels of expression of nucleotidase are correlated with poor outcome in patients with AML (4). On the other hand, nucleotide pool imbalance caused by increased or decreased nucleotidase activity is also known to cause disorders. For instance, 5′-nucleotidase deficiency has been implicated in hereditary hemolytic anemia (16). Here, we report that GNN/TU12B1-TY, a putative ATP-regulated 5′-nucleotidase is overexpressed in two subtypes of human AML. In addition, our data implicate that *Gnn* is a target of the retroviral insertion in the cVIS *Evi12*. Together, this strongly suggests that *Gnn* is a proto-oncogene that may be involved in the progression of certain types of leukemia. Moreover, since *Evi12* is also found as a common insertion in lymphoma mouse models, it would be interesting to determine *Gnn* expression in human lymphoma as well.

ACKNOWLEDGMENTS

We thank Neal Copeland (NCI, Frederick, Md.) for providing the AKXD tumors, A. van Hoven-Beijen and C. Erpelinck for technical assistance, L. H. J. Aarts for help with confocal laser scanning microscopy, and R. Verhaak and M. Bijl for analysis of the Affymetrix GeneChip data with the Micro Array Data Explorer (MADEx).

This work was supported in part by grant NKB-EUR 2000-2007 from the Dutch Cancer Society (KWF).

REFERENCES

1. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276–280.
2. **Bianchi, V., and J. Spychala.** 2003. Mammalian 5′-nucleotidases. J. Biol. Chem. **278:**46195–46198.
3. **Erkeland, S. J., M. Valkhof, C. Heijmans-Antonissen, A.van Hoven-Beijen, R. Delwel, M. H. A. Hermans, and I. P. Touw.** 2004. Large-scale identification of disease genes involved in acute myeloid leukemia. J. Virol. **78:**1971–1980.
4. **Galmarini, C. M., K. Graham, X. Thomas, F. Calvo, P. Rousselot, A. El Jafaari, E. Cros, J. R. Mackey, and C. Dumontet.** 2001. Expression of high Km 5′-nucleotidase in leukemic blasts is an independent prognostic factor in adults with acute myeloid leukemia. Blood **98:**1922–1926.
5. **Gattiker, A., E. Gasteiger, and A. Bairoch.** 2002. ScanProsite: a reference implementation of a PROSITE scanning tool. Appl. Bioinformatics **1:**107–108.
6. **Geer, L. Y., M. Domrachev, D. J. Lipman, and S. H. Bryant.** 2002. CDART: protein homology by domain architecture. Genome Res. **12:**1619–1623.
7. **Jonkers, J., and A. Berns.** 1996. Retroviral insertional mutagenesis as a strategy to identify cancer genes. Biochim. Biophys. Acta **1287:**29–57.
8. **Joosten, M., P. J. Valk, Y. Vankan, N. de Both, B. Lowenberg, and R. Delwel.** 2000. Phenotyping of Evi1, Evi11/Cb2, and Evi12 transformed leukemias isolated from a novel panel of cas-Br-M murine leukemia virus-infected mice. Virology **268:**308–318.
9. **Joosten, M., Y. Vankan-Berkhoudt, M. Tas, M. Lunghi, Y. Jenniskens, E. Parganas, P. J. Valk, B. Lowenberg, E. van den Akker, and R. Delwel.** 2002. Large-scale identification of novel potential disease loci in mouse leukemia applying an improved strategy for cloning common virus integration sites. Oncogene **21:**7247–7255.
10. **Li, J., H. Shen, K. L. Himmel, A. J. Dupuy, D. A. Largaespada, T. Nakamura, J. D. Shaughnessy, Jr., N. A. Jenkins, and N. G. Copeland.** 1999. Leukaemia disease genes: large-scale cloning and pathway predictions. Nat. Genet. **23:** 348–353.
11. **Lund, A. H., G. Turner, A. Trubetskoy, E. Verhoeven, E. Wientjens, D. Hulsman, R. Russell, R. A. DePinho, J. Lenz, and M. van Lohuizen.** 2002. Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice. Nat. Genet. **32:**160–165.
12. **Lupas, A. N., and J. Martin.** 2002. AAA proteins. Curr. Opin. Struct. Biol. **12:**746–753.
13. **Mikkers, H., J. Allen, P. Knipscheer, L. Romeijn, A. Hart, E. Vink, and A. Berns.** 2002. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nat. Genet. **32:**153–159.
14. **Ofir, R., J. Gopas, E. Aflalo, and Y. Weinstein.** 1992. Oligoclonality of Moloney leukemias. Leuk. Res. **16:**797–806.
15. **Oka, J., A. Matsumoto, Y. Hosokawa, and S. Inoue.** 1994. Molecular cloning of human cytosolic purine 5′-nucleotidase. Biochem. Biophys. Res. Commun. **205:**917–922.
16. **Rees, D. C., J. A. Duley, and A. M. Marinaki.** 2003. Pyrimidine 5′ nucleotidase deficiency. Br. J. Haematol. **120:**375–383.
17. **Suzuki, T., H. Shen, K. Akagi, H. C. Morse, J. D. Malley, D. Q. Naiman, N. A. Jenkins, and N. G. Copeland.** 2002. New genes involved in cancer identified by retroviral tagging. Nat. Genet. **32:**166–174.
18. **Tusher, V. G., R. Tibshirani, and G. Chu.** 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. USA **98:**5116–5121.
19. **Valk, P. J. M., S. Hol, Y. Vankan, J. N. Ihle, D. Askew, N. A. Jenkins, D. J. Gilbert, N. G. Copeland, N. J. de Both, B. Löwenberg, and R. Delwel.** 1997. The genes encoding the peripheral cannabinoid receptor and α-L-fucosidase are located near a newly identified common virus integration site, *Evi11*. J. Virol. **71:**6796–6804.
20. **Valk, P. J. M., Y. Vankan, M. Joosten, N. A. Jenkins, N. G. Copeland, B. Löwenberg, and R. Delwel.** 1999. Retroviral insertions in *Evi12*, a novel common virus integration site upstream of *Tra1*/*Grp94*, frequently coincide with insertions in the gene encoding the peripheral cannabinoid receptor *Cnr2*. J. Virol. **73:**3595–3602.
21. **Valk, P. J., R. G. Verhaak, M. A. Beijen, C. A. Erpelinck, S. Barjesteh van Waalwijk van Doorn-Khosrovani, J. M. Boer, H. B. Beverloo, M. J. Moorhouse, P. J. van der Spek, B. Lowenberg, and R. Delwel.** 2004. Prognostically useful gene-expression profiles in acute myeloid leukemia. N. Engl. J. Med. **350:**1617–1628.
22. **van de Geijn, G. J., J. Gits, L. H. Aarts, C. Heijmans-Antonissen, and I. P. Touw.** 2004. G-CSF receptor truncations found in SCN/AML relieve SOCS3-controlled inhibition of STAT5 but leave suppression of STAT3 intact. Blood **104:**667–674.